Computers and Mathematics with Applications 62 (2011) 1957–1968

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

On numerical stabilization in the solution of Saint-Venant equations using the finite element method

Fatemeh Zarmehi [a], Ali Tavakoli [a,*], Majid Rahimpour [b]

[a] Department of Mathematics, Vali-e-Asr University of Rafsanjan, Iran
[b] Department of Water Engineering, Shahid Bahonar University of Kerman, Iran

ARTICLE INFO

Article history:
Received 5 December 2010
Received in revised form 19 June 2011
Accepted 20 June 2011

Keywords:
Hyperbolic partial differential equation
Saint-Venant equations
Finite element method
M-matrix

ABSTRACT

Solving the Saint-Venant equations by using numerical schemes like finite difference and finite element methods leads to some unwanted oscillations in the water surface elevation. The reason for these oscillations lies in the method used for the approximation of the nonlinear terms. One of the ways of smoothing these oscillations is by adding artificial viscosity into the scheme. In this paper, by using a suitable discretization, we first solve the one-dimensional Saint-Venant equations by a finite element method and eliminate the unwanted oscillations without using an artificial viscosity. Second, our main discussion is concentrated on numerical stabilization of the solution in detail. In fact, we first convert the systems resulting from the discretization to systems relating to just water surface elevation. Then, by using M-matrix properties, the stability of the solution is shown. Finally, two numerical examples of critical and subcritical flows are given to support our results.

1. Introduction

Unsteady flow is of great interest to hydraulic engineers. Such flows can be described by the Saint-Venant equations which consist of the conservation of mass and momentum equations. The Saint-Venant equations are also nonlinear hyperbolic partial differential equations. However, a general closed form solution of these equations is not available, except for certain special simplified conditions and they must be solved using an appropriate numerical technique [1]. In typical hydraulics textbooks (e.g. [2,3]) these equations are derived from the incompressible Navier–Stokes equations. Over the past few years, a wide range of numerical schemes based on the finite difference [4], finite element [5,6], and finite volume [7] methods have been applied to the open channel flow equations.

In [8], Kiladze has studied the stability of finite difference schemes for solving Saint-Venant equations. Also in [9], Bastin et al. have investigated the issue of the exponential stability (in L_2-norm) of the classical solutions of the linearized Saint-Venant equations for a sloping channel. Furthermore, in [10], Thual et al. have derived and displayed the stability of the homogeneous and steady flow on the basis of the one-dimensional Saint-Venant equations for free surface and shallow water flows of constant slope through graphs.

In this paper we consider an initial–boundary value Saint-Venant problem for unsteady flow in an open channel having no lateral inflow or outflow for one dimension as follows:

* Corresponding author. Fax: +98 391 3202270.
E-mail addresses: f.zarmehi@mail.vru.ac.ir (F. Zarmehi), tavakoli@mail.vru.ac.ir (A. Tavakoli), rahimpour@mail.uk.ac.ir (M. Rahimpour).

0898-1221/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2011.06.039

$$\begin{cases} \dfrac{\partial Q}{\partial t} + \dfrac{\partial}{\partial x}\left(\dfrac{Q^2}{A}\right) + gA\dfrac{\partial h}{\partial x} + \dfrac{gn^2|Q|Q}{R^{4/3}A} = 0 & \text{momentum equation,} \\ \dfrac{\partial h}{\partial t} + \dfrac{1}{B}\dfrac{\partial Q}{\partial x} = 0 & \text{continuity equation,} \\ Q(x,0) = Q^0 & 0 \le x < L, \\ h(x,0) = h^0 & 0 \le x \le L, \\ Q(L,t) = 0 & t \ge 0, \\ h(0,t) = h_0 & t > 0, \end{cases} \tag{1}$$

in which x = distance along the channel length, t = time, A = flow area, B = top water surface width, g = acceleration due to gravity, Q = discharge, h = water surface elevation, R = hydraulic radius, n = Manning coefficient, and L = length of channel, and also h^0, h_0 and Q^0 are positive constant scalars. In general A and R are functions of h (i.e. $A = A(h)$, $R = R(h)$).

We first discretize the Saint-Venant equations by the finite element method and then prove the numerical stabilization of the method in detail. In Section 2, the finite element discretizations of the Saint-Venant equations are given. In Section 3, by using the properties of the M-matrix, the stability of the water surface elevation h in the Saint-Venant equations is shown. Two numerical results are given in Section 4 to support our theoretical discussion. Finally, in Section 5 the conclusion and some comments as regards future work are given.

2. Discretization of the Saint-Venant equations

In this section, we explain the process of linearization and determine the shape functions of the finite element method for Saint-Venant equations.

2.1. Linearity

In order to present the variational form of the Saint-Venant equations, we focus our attention on the discretization with respect to the time variable. Thus, we choose a positive integer N, and let Δt denote the corresponding time step: $\Delta t = T/N$, and t_n the subdivisions of $[0, T]$:

$$t_n = n\Delta t, \quad 0 \le n \le N.$$

For linearity, we consider the terms of Saint-Venant equations as follows:

$$\begin{cases} \dfrac{\partial Q(x,t_n)}{\partial t} + \dfrac{\partial}{\partial x}\left(\dfrac{Q^2(x,t_{n+1})}{A(x,t_n)}\right) + gA(x,t_n)\dfrac{\partial h(x,t_{n+1})}{\partial x} + \dfrac{gn^2|Q(x,t_n)|Q(x,t_{n+1})}{R^{4/3}(x,t_n)A(x,t_n)} = 0, \\ \dfrac{\partial h(x,t_n)}{\partial t} + \dfrac{1}{B}\dfrac{\partial Q(x,t_{n+1})}{\partial x} = 0. \end{cases} \tag{2}$$

Now, by Taylor expansion we get

$$\begin{cases} \text{(a)} \ \dfrac{\partial Q(x,t_n)}{\partial t} \cong \dfrac{Q(x,t_{n+1}) - Q(x,t_n)}{\Delta t}, \\ \text{(b)} \ \dfrac{\partial h(x,t_n)}{\partial t} \cong \dfrac{h(x,t_{n+1}) - h(x,t_n)}{\Delta t}. \end{cases} \tag{3}$$

Moreover,

$$\begin{aligned} Q^2(x,t_{n+1}) &\cong Q^2(x,t_n) + \Delta t\frac{\partial Q^2(x,t_n)}{\partial t} \\ &\cong Q^2(x,t_n) + 2\Delta t Q(x,t_n)\frac{\partial Q(x,t_n)}{\partial t}. \end{aligned} \tag{4}$$

Then, by substituting Eq. (3)(a) in Eq. (4) we obtain

$$Q^2(x,t_{n+1}) \cong -Q^2(x,t_n) + 2Q(x,t_n)Q(x,t_{n+1}). \tag{5}$$

Substituting Eqs. (3) and (5) into Eqs. (2) and simplifying, we can write the discrete form of (1) as follows:

$$\begin{cases} \dfrac{1}{\Delta t}Q(x,t_{n+1}) + \dfrac{\partial}{\partial x}\left(\dfrac{2Q(x,t_n)Q(x,t_{n+1})}{A(x,t_n)}\right) + gA(x,t_n)\dfrac{\partial h(x,t_{n+1})}{\partial x} \\ \quad + \dfrac{gn^2|Q(x,t_n)|Q(x,t_{n+1})}{R^{4/3}(x,t_n)A(x,t_n)} = \dfrac{1}{\Delta t}Q(x,t_n) + \dfrac{\partial}{\partial x}\left(\dfrac{Q^2(x,t_n)}{A(x,t_n)}\right), \\ \dfrac{1}{\Delta t}h(x,t_{n+1}) + \dfrac{1}{B}\dfrac{\partial Q(x,t_{n+1})}{\partial x} = \dfrac{1}{\Delta t}h(x,t_n), \\ Q(x,0) = Q^0, \quad 0 \le x < L, \\ h(x,0) = h^0, \quad 0 \le x \le L, \\ Q(L,t_n) = 0, \quad n = 0,\ldots,N, \\ h(0,t_n) = h_0, \quad n = 1,\ldots,N. \end{cases} \tag{6}$$

2.2. The variational weak form and shape functions

In this subsection, we present the finite element approximation to problem (6). The variational form of problem (6) is that of finding $Q(x,t_{n+1}) \in V = \{Q(x,t_k) \in H^1(\Omega) : Q(L,t_k) = 0, k = 0,\ldots,N\}$, and $h(x,t_{n+1}) \in H = \{h(x,t_k) \in H^1(\Omega) : h(0,t_k) = h_0,\ k = 1,\ldots,N\}$ such that

$$\begin{aligned} &d(h,v) + m(Q,v) + b(Q,v) = (\alpha,v)_0 \quad \forall v \in H, \\ &s(h,e) + w(Q,e) = (\beta,e)_0 \qquad\qquad \forall e \in V, \end{aligned} \tag{7}$$

where $\Omega = [0,L]$, $(.,.)_0$ is an inner product in the $L_2(\Omega)$ space, and the bilinear forms on $V \times H$ are given respectively by

$$\begin{aligned} m(Q,v) &= \int_\Omega \left(\frac{1}{\Delta t} + \frac{gn^2|Q(x,t_n)|}{R^{4/3}(x,t_n)A(x,t_n)}\right) Q(x,t_{n+1})v\,dx, \\ b(Q,v) &= -2\int_\Omega \frac{Q(x,t_n)}{A(x,t_n)}Q(x,t_{n+1})v'\,dx + \frac{2Q(x,t_n)Q(x,t_{n+1})}{A(x,t_n)}v\Big|_{\partial\Omega}, \\ d(h,v) &= -g\int_\Omega h(x,t_{n+1})(A(x,t_n)v)'\,dx + gA(x,t_n)h(x,t_{n+1})v\Big|_{\partial\Omega}, \\ s(h,e) &= \frac{1}{\Delta t}\int_\Omega h(x,t_{n+1})e\,dx, \\ w(Q,e) &= \frac{-1}{B}\int_\Omega Q(x,t_{n+1})e'\,dx + \frac{1}{B}Q(x,t_{n+1})e\Big|_{\partial\Omega}, \\ (\alpha,v) &= \int_\Omega \alpha v\,dx, \end{aligned} \tag{8}$$

where $\partial\Omega$ is the boundary of Ω and $v|_{\partial\Omega}$ is the restriction of v on $\partial\Omega$.

For approximated discrete mixed formulation of (7), we choose a positive integer M, and let Δx denote the corresponding displacement step: $\Delta x = L/M$, and x_i the subdivisions of $[0,L]$:

$$x_i = i\Delta x \quad 0 \le i \le M.$$

Now, suppose that $V_a = span\{\varphi_i\}_{i=0}^{M-1} \subset V$, and that $H_a = span\{\psi_j\}_{j=1}^{M} \subset H$ where the shape functions φ_i and ψ_j are piecewise linear polynomial functions on the nodes $x_0,\ldots,x_{M-1}$ and $x_1,\ldots,x_M$, respectively (Fig. 1). For example, the shape function φ_i is defined by

$$\varphi_i(x) = \begin{cases} \dfrac{x_{i+1}-x}{x_{i+1}-x_i}, & x \in \Omega_{i+1} \\ \dfrac{x-x_{i-1}}{x_i-x_{i-1}}, & x \in \Omega_i \\ 0, & \text{otherwise.} \end{cases}$$

We consider the variational problem of finding $(Q_a(x,t_{n+1}), h_a(x,t_{n+1})) \in V_a \times H_a$ such that

$$\begin{aligned} &d(h_a,v_a) + m(Q_a,v_a) + b(Q_a,v_a) = (\alpha,v_a)_0 \quad \forall v_a \in H_a, \\ &s(h_a,e_a) + w(Q_a,e_a) = (\beta,e_a)_0 \qquad\qquad \forall e_a \in V_a. \end{aligned} \tag{9}$$

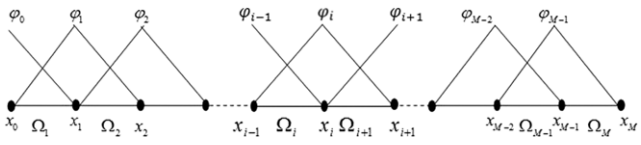

Fig. 1. Schematic representation of the nodal points and basis functions.

Remark 1. Since the value of h is nonzero in the first channel (i.e. the constant h_0), then we can consider

$$h_a(x, t_{n+1}) = h_0\psi_0 + \sum_{j=1}^{M} \beta_j \psi_j,$$

where $\beta_j, j = 1, \ldots, M$, are unknown scalar parameters and ψ_0 is a piecewise linear polynomial function at the node $x = 0$. We note that this one additional shape function does not change the dimension of the stiffness matrix.

Remark 2. In [11], Granatowicz and Szymkiewicz have used averaging of the nonlinear terms in space to eliminate oscillations. Here, we use a rather different way, namely, the averaging of $h(x, t_n)$ and $Q(x, t_n)$ in space for time $t = t_{n+1}$. In other words, if $h(x, t_n) = \sum_{j=1}^{M} \beta_j\psi_j$ and $Q(x, t_n) = \sum_{i=0}^{M-1} \alpha_i\varphi_i$, then for time $t = t_{n+1}$ we make the replacements $h(x, t_n) = \sum_{j=1}^{M-1} \frac{(\beta_j+\beta_{j+1})}{2}\psi_j + \beta_M\psi_M$ and $Q(x, t_n) = \sum_{i=0}^{M-2} \frac{(\alpha_i+\alpha_{i+1})}{2}\varphi_i + \alpha_{M-1}\varphi_{M-1}$.

3. Stability of the solution

Solving the Saint-Venant equations by the finite element method leads to some unwanted oscillations in the water surface elevation h. This yields an instability of the solution, for which the main reason is related to the method used for the approximation of the nonlinear terms. In this section, we explain why the water surface elevation h is stable through the discretization problem (9). To this end, one can use the M-matrix properties to show the stability of h. Of course, we need a condition – called the stability condition – to show that h is stable. Already, many researchers have used the M-matrix properties to prove the numerical stability (for instance, see [12] and [13] Chapter 2).

Definition 1. A matrix $A = (a_{i,j})_{1\le i,j\le n}$ is said to be an M-matrix if $a_{i,j} \le 0$ for $i \ne j$, and the real parts of the eigenvalues are positive.

In the following, for simplicity we abbreviate $U(x_i, t_j)$ as U_i^j. Also, let v_i^j be the velocity at the point (x_i, t_j). In order to show the stability of h, we need the following assumptions:

A1: $\mathbf{F}_r \le 1$, where $\mathbf{F}_r = \frac{v^0}{\sqrt{gh^0}}$ is a Froude number and $v^0 = \frac{Q^0}{A^0}$ is the velocity at t_0.

A2: For $i = 0, \ldots, M$ and $j = 1, \ldots, N$, we have

$$|v^0| \ge |v_i^{j-1}|, \qquad h^0 \le h_i^{j-1}.$$

In fact by assumption A1, we consider that there is a critical or subcritical flow only at t_0. The assumption A2 is valid for many open channel problems. For instance, when a sluice gate at the downstream end of an open channel is suddenly closed at time $t = 0$, the initial and boundary conditions of problem (1) with assumption A2 would be valid (see Examples 1 and 2). Also, when a sluice gate at the upstream end of an open channel is suddenly opened at time $t = 0$, the initial–boundary conditions of problem (1) should be changed and thus assumption A2 would be changed, too. Hence, in this case, the stability of the solution should be verified with the new conditions.

The general matrix form of the system resulting from (9) at time $t = t_j$ reads

$$\begin{bmatrix} D & E \\ S & W \end{bmatrix} \begin{bmatrix} \mathbf{h}^j \\ \mathbf{Q}^j \end{bmatrix} = \begin{bmatrix} F^j \\ G^j \end{bmatrix}, \tag{10}$$

in which $\mathbf{h}^j = [h_1^j, h_2^j, \ldots, h_M^j]^T$, $\mathbf{Q}^j = [Q_0^j, Q_1^j, \ldots, Q_{M-1}^j]^T$, and $F^j = [F_1^j, F_2^j, \ldots, F_M^j]^T$, $G^j = [G_0^j, G_1^j, \ldots, G_{M-1}^j]^T$ where $F_k^j = (\alpha, \psi_k), k = 1, \ldots, M$, and $G_k^j = (\beta, \varphi_k), k = 0, \ldots, M-1$, and also the entries of the matrices D, E, S and W are respectively defined as follows:

$$\begin{aligned}
&D_{ij} = d(\psi_i, \psi_j), && i, j = 1, \ldots, M,\\
&E_{ij} = m(\varphi_i, \psi_j) + b(\varphi_i, \psi_j), && i = 0, \ldots, M-1,\ j = 1, \ldots, M,\\
&S_{ij} = s(\psi_i, \varphi_j), && i = 1, \ldots, M,\ j = 0, \ldots, M-1,\\
&W_{ij} = w(\varphi_i, \varphi_j), && i, j = 0, \ldots, M-1.
\end{aligned}$$

Suppose that $a = \frac{\Delta x}{6\Delta t} - \frac{Q^0}{A^0} + \frac{gn^2|Q^0|\Delta x}{6(R^0)^{4/3}A^0}$, $b = \frac{2\Delta x}{3\Delta t} + \frac{2gn^2|Q^0|\Delta x}{3(R^0)^{4/3}A^0}$ and $c = \frac{\Delta x}{6\Delta t} + \frac{Q^0}{A^0} + \frac{gn^2|Q^0|\Delta x}{6(R^0)^{4/3}A^0}$. Also, let the constant values A^0, Q^0 and R^0 represent the values of A, Q and R at time $t_0 = 0$, respectively, where $A^0 = A(h^0)$ and $R^0 = R(h^0)$. It is readily seen that for $t_1 = \Delta t$, the general forms of the blocks of the stiffness matrix are given by

$$D = \begin{bmatrix} 0 & \frac{gA^0}{2} & & & & & \\ \frac{-gA^0}{2} & 0 & \frac{gA^0}{2} & & & & \\ & \frac{-gA^0}{2} & 0 & \frac{gA^0}{2} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \frac{-gA^0}{2} & 0 & \frac{gA^0}{2} & \\ & & & & \frac{-gA^0}{2} & \frac{gA^0}{2} & \end{bmatrix},$$

$$E = \begin{bmatrix} a & b & c & & & & \\ & a & b & c & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & a & b & c & \\ & & & & a & b & \\ & & & & & a & \end{bmatrix},$$

$$S = \begin{bmatrix} \frac{\Delta x}{6\Delta t} & & & & & & \\ \frac{2\Delta x}{3\Delta t} & \frac{\Delta x}{6\Delta t} & & & & & \\ \frac{\Delta x}{6\Delta t} & \frac{2\Delta x}{3\Delta t} & \frac{\Delta x}{6\Delta t} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \frac{\Delta x}{6\Delta t} & \frac{2\Delta x}{3\Delta t} & \frac{\Delta x}{6\Delta t} & & \end{bmatrix},$$

$$W = \begin{bmatrix} \frac{-1}{2B} & \frac{1}{2B} & & & & & \\ \frac{-1}{2B} & 0 & \frac{1}{2B} & & & & \\ & \frac{-1}{2B} & 0 & \frac{1}{2B} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \frac{-1}{2B} & 0 & \frac{1}{2B} & \\ & & & & \frac{-1}{2B} & 0 & \end{bmatrix}.$$

In order to show that the solution h is stable, we proceed as follows.

For the moment, we can assume that h_M^1 and Q_{M-1}^1 are known (we note that since $Q(L, t_n) = 0$, $n = 0, \ldots, N$, we hence have $Q_{M-1}^1 \cong 0$ and $h_M^1 = h_{max}$). Therefore, by this assumption, we cancel the first rows of D, E, S and W. Also, the last columns of D, E, S and W are transformed to the right hand side ones. Hence, the system (10) is converted to the following system:

$$\begin{bmatrix} D^r & E^r \\ S^r & W^r \end{bmatrix} \begin{bmatrix} \mathbf{h}^{1,r} \\ \mathbf{Q}^{1,r} \end{bmatrix} = \begin{bmatrix} F^{1,r} \\ G^{1,r} \end{bmatrix}, \tag{11}$$

where

$$D^r = \begin{bmatrix} \frac{-gA^0}{2} & 0 & \frac{gA^0}{2} & & & & \\ & \frac{-gA^0}{2} & 0 & \frac{gA^0}{2} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \frac{-gA^0}{2} & 0 & \frac{gA^0}{2} & \\ & & & & \frac{-gA^0}{2} & 0 & \\ & & & & & \frac{-gA^0}{2} & \end{bmatrix},$$

$$E^r = \begin{bmatrix} 0 & a & b & c & & & \\ & 0 & a & b & c & & \\ & & \ddots & \ddots & \ddots & \ddots & \\ & & & 0 & a & b & c \\ & & & & 0 & a & b \\ & & & & & 0 & a \end{bmatrix},$$

$$S^r = \begin{bmatrix} \frac{2\Delta x}{3\Delta t} & \frac{\Delta x}{6\Delta t} & & & & \\ \frac{\Delta x}{6\Delta t} & \frac{2\Delta x}{3\Delta t} & \frac{\Delta x}{6\Delta t} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \frac{\Delta x}{6\Delta t} & \frac{2\Delta x}{3\Delta t} & \frac{\Delta x}{6\Delta t} & \\ & & & \frac{\Delta x}{6\Delta t} & \frac{2\Delta x}{3\Delta t} & \end{bmatrix},$$

$$W^r = \begin{bmatrix} \frac{-1}{2B} & 0 & \frac{1}{2B} & & & \\ & \frac{-1}{2B} & 0 & \frac{1}{2B} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \frac{-1}{2B} & 0 & \frac{1}{2B} \\ & & & & \frac{-1}{2B} & 0 \\ & & & & & \frac{-1}{2B} \end{bmatrix},$$

$$F^{1,r} = \begin{bmatrix} F_2^1 \\ F_3^1 \\ \vdots \\ F_{M-2}^1 - cQ_{M-1}^1 \\ F_{M-1}^1 - \frac{gA^0}{2}h_M^1 - bQ_{M-1}^1 \\ F_M^1 - \frac{gA^0}{2}h_M^1 - aQ_{M-1}^1 \end{bmatrix},$$

$$G^{1,r} = \begin{bmatrix} G_1^1 \\ G_2^1 \\ \vdots \\ G_{M-3}^1 \\ G_{M-2}^1 - \frac{1}{2B}Q_{M-1}^1 \\ G_{M-1}^1 - \frac{\Delta x}{6\Delta t}h_M^1 \end{bmatrix},$$

$$\mathbf{h}^{1,r} = \begin{bmatrix} h_1^1 & h_2^1 & \dots & h_{M-1}^1 \end{bmatrix}^T, \qquad \mathbf{Q}^{1,r} = \begin{bmatrix} Q_0^1 & Q_1^1 & \dots & Q_{M-2}^1 \end{bmatrix}^T.$$

The two equations removed are as follows:

$$\begin{aligned} &\frac{gA^0}{2}h_2^1 + aQ_0^1 + bQ_1^1 + cQ_2^1 = F_1^1, \\ &\frac{\Delta x}{6\Delta t}h_1^1 - \frac{1}{2B}Q_0^1 + \frac{1}{2B}Q_1^1 = G_0^1. \end{aligned} \tag{12}$$

The simplified system (11) is as follows:

$$\begin{aligned} &\text{(a) } D^r\mathbf{h}^{1,r} = F^{1,r} - E^r\mathbf{Q}^{1,r}, \\ &\text{(b) } W^r\mathbf{Q}^{1,r} = G^{1,r} - S^r\mathbf{h}^{1,r}. \end{aligned} \tag{13}$$

Since W^r is an upper triangular matrix and all diagonal entries are nonzero, then W^r is nonsingular. Therefore by deriving $\mathbf{Q}^{1,r}$ from (13)(b) and substituting it in (13)(a), we obtain

$$(E^r(W^r)^{-1}S^r - D^r)\mathbf{h}^{1,r} = E^r(W^r)^{-1}G^{1,r} - F^{1,r}. \tag{14}$$

In order to show that the solution h is stable at t_1, we should prove that $E^r(W^r)^{-1}S^r - D^r$ is an M-matrix. To this end, we first note that $(W^r)^{-1}S^r$ is a nonpositive matrix and an upper Hessenberg matrix (i.e. $((W^r)^{-1}S^r)_{ij} = 0$ for all $i > j+1$ and $((W^r)^{-1}S^r)_{ij} \le 0$ for all $i \le j+1$), since $(W^r)^{-1}$ is an upper triangular nonpositive matrix, and S^r is a tri-diagonal nonnegative matrix. On the other hand, since E^r is an upper triangular matrix with diagonal elements zero, then $E^r(W^r)^{-1}S^r$ is an upper triangular matrix, too.

Now, we find a condition under which the matrix $E^r(W^r)^{-1}S^r - D^r$ is an M-matrix.

$E^r(W^r)^{-1}S^r - D^r$ is clearly an upper triangular matrix and thus its eigenvalues are diagonal entries. If the off-diagonal entries of $E^r(W^r)^{-1}S^r - D^r$ are nonpositive and diagonal positive, then by definition this matrix would be an M-matrix. This would happen if the absolute values of the diagonal entries $E^r(W^r)^{-1}S^r$ were less than the corresponding diagonal entries of $(-D^r)$, which implies

$$diag(E^r(W^r)^{-1}S^r - D^r) > 0.$$

This means that we need the following condition:

$$\mu^1 \le \frac{\Delta x}{\Delta t} < \omega^1 \tag{15}$$

where

$$\mu^1 = \frac{3Q^0}{A^0} - \frac{gn^2|Q^0|\Delta x}{(R^0)^{4/3}A^0}, \tag{16}$$

and

$$\omega^1 = 3\left(\frac{Q^0}{A^0} - \frac{gn^2|Q^0|\Delta x}{6(R^0)^{4/3}A^0}\right) + 3\sqrt{\left(\frac{Q^0}{A^0} - \frac{gn^2|Q^0|\Delta x}{6(R^0)^{4/3}A^0}\right)^2 + gh^0}. \tag{17}$$

This condition is derived from the following inequalities:

$$\frac{4\Delta x}{3\Delta t} - \frac{4Q^0}{A^0} + \frac{4gn^2|Q^0|\Delta x}{3(R^0)^{4/3}A^0} \ge 0,$$

or

$$\frac{\Delta x}{\Delta t} \ge \frac{3Q^0}{A^0} - \frac{gn^2|Q^0|\Delta x}{(R^0)^{4/3}A^0}, \tag{18}$$

and

$$\frac{-B\Delta x}{3\Delta t}\left(\frac{\Delta x}{6\Delta t} - \frac{Q^0}{A^0} + \frac{gn^2|Q^0|\Delta x}{6(R^0)^{4/3}A^0}\right) + \frac{gA^0}{2} > 0,$$

or

$$\frac{B\Delta x}{3\Delta t}\left(\frac{\Delta x}{6\Delta t} - \frac{Q^0}{A^0} + \frac{gn^2|Q^0|\Delta x}{6(R^0)^{4/3}A^0}\right) < \frac{gA^0}{2}. \tag{19}$$

We note that the Manning coefficient is very small. Hence, for a sufficiently small Δx, we have $0 < \mu^1 < \omega^1$.

Remark 3. Since the Manning coefficient is very small, for computational purposes with an sufficiently small Δx, one can eliminate the term $\frac{gn^2|Q^0|\Delta x}{(R^0)^{4/3}A^0}$ from μ^1 and ω^1. In other words, we can discard the term $\frac{gn^2|Q^0|\Delta x}{(R^0)^{4/3}A^0}$ if it is less than a very small threshold value α. Then, Δx should be chosen less than $\frac{\alpha(R^0)^{4/3}A^0}{gn^2|Q^0|}$. Hence, one can consider the stability condition as

$$3\frac{Q^0}{A^0} \le \frac{\Delta x}{\Delta t} < 3\frac{Q^0}{A^0} + 3\sqrt{\left(\frac{Q^0}{A^0}\right)^2 + gh^0}, \tag{20}$$

where Δx is sufficiently small.

Now, we discuss the numerical stability of h for the next times. Let $E^{j,r}$, $W^{j,r}$, $S^{j,r}$ and $D^{j,r}$ be the corresponding matrices of E^r, W^r, S^r and D^r at time t_j, respectively. We will show that the matrix $E^{j,r}(W^{j,r})^{-1}S^{j,r} - D^{j,r}$ remains an M-matrix for $t = t_j, j = 2, \ldots, N$. We note that the matrices $W^{j,r}$ and $S^{j,r}$ do not change, since they do not depend on $h(x, t_{j-1})$ and $Q(x, t_{j-1})$. As the water surface elevation h changes with time, the first upper diagonal in $D^{j,r}$ for $j = 2, \ldots, N$ would not remain zero. Now, we show that $D^{j,r}_{i,i+1} > 0$ for $i = 1, \ldots, M-1$. For this purpose, we consider ψ_i to be the shape function at the node x_i. Then, we can write

$$\begin{aligned}
D^{j,r}_{i,i+1} &= -gB\int_\Omega \psi_i \left(h(x, t_j)\psi_i\right)' \mathrm{d}x \\
&= -gB\int_{\Omega_i \cup \Omega_{i+1}} \psi_i \left(h(x, t_j)\psi_i\right)' \mathrm{d}x \\
&= -gB\left(\int_{\Omega_i} \psi_i \left((\alpha_{i-1}\psi_{i-1} + \alpha_i\psi_i)\,\psi_i\right)' \mathrm{d}x + \int_{\Omega_{i+1}} \psi_i \left((\alpha_i\psi_i + \alpha_{i+1}\psi_{i+1})\,\psi_i\right)' \mathrm{d}x\right) \\
&= gB\left(\frac{\alpha_{i-1} + \alpha_{i+1}}{2}\right) > 0,
\end{aligned}$$

where α_{i-1}, α_i and α_{i+1} are the computed values of h at the nodes x_{i-1}, x_i and x_{i+1}, at $t = t_{j-1}$, respectively. We note that α_k, $k = i-1, i, i+1$, are nonnegative, since $h^j_k > 0$, for any j, k.

Similarly for the other times the matrix $E^{j,r}(W^{j,r})^{-1}S^{j,r} - D^{j,r}$ is an M-matrix if

$$\frac{\Delta x}{\Delta t} \ge \frac{3Q^{j-1}_i}{A^{j-1}_i} - \frac{gn^2|Q^{j-1}_i|\Delta x}{(R^{j-1}_i)^{4/3}A^{j-1}_i}, \tag{21}$$

and

$$\frac{B\Delta x}{3\Delta t}\left(\frac{\Delta x}{6\Delta t} - \frac{Q^{j-1}_i}{A^{j-1}_i} + \frac{gn^2|Q^{j-1}_i|\Delta x}{6(R^{j-1}_i)^{4/3}A^{j-1}_i}\right) < \frac{gA^{j-1}_i}{2}, \tag{22}$$

hold. On disregarding the very small term $\frac{gn^2|Q^{j-1}_i|\Delta x}{(R^{j-1}_i)^{4/3}A^{j-1}_i}$, the conditions (21) and (22) reduce to the following:

$$\frac{\Delta x}{\Delta t} \ge \frac{3Q^{j-1}_i}{A^{j-1}_i}, \tag{23}$$

and

$$\frac{B\Delta x}{3\Delta t}\left(\frac{\Delta x}{6\Delta t} - \frac{Q^{j-1}_i}{A^{j-1}_i}\right) < \frac{gA^{j-1}_i}{2}. \tag{24}$$

By assumption A2, $\frac{3Q^0}{A^0} \ge \frac{3Q^{j-1}_i}{A^{j-1}_i}$ is satisfied. Hence, the condition (20) is a sufficient condition for (23) to hold. It remains to show that the condition (24) is satisfied. To this end, by the simplicity of condition (24), we have

$$3\frac{Q^{j-1}_i}{A^{j-1}_i} - 3\sqrt{\left(\frac{Q^{j-1}_i}{A^{j-1}_i}\right)^2 + gh^{j-1}_i} < \frac{\Delta x}{\Delta t} < 3\frac{Q^{j-1}_i}{A^{j-1}_i} + 3\sqrt{\left(\frac{Q^{j-1}_i}{A^{j-1}_i}\right)^2 + gh^{j-1}_i}. \tag{25}$$

The left hand side of the above inequality holds obviously. On the other hand,

$$3\frac{Q^{j-1}_i}{A^{j-1}_i} + 3\sqrt{\left(\frac{Q^{j-1}_i}{A^{j-1}_i}\right)^2 + gh^{j-1}_i} \ge 3\sqrt{gh^0},$$

is satisfied by assumption A2. Therefore, the condition (25) holds if

$$0 < \frac{\Delta x}{\Delta t} \leq 3\sqrt{gh^0}. \tag{26}$$

Furthermore, by assumption A1 we have

$$\frac{Q^0}{A^0} \leq \sqrt{gh^0}. \tag{27}$$

Finally, by (20), (26) and (27) the stability condition for any time reads

$$3\frac{Q^0}{A^0} \leq \frac{\Delta x}{\Delta t} \leq 3\sqrt{gh^0}. \tag{28}$$

Remark 4. We note that by the stability condition (28) and assumption A2, we have

$$\left| v^0 \frac{\Delta t}{\Delta x} \right| = \left| \frac{Q^0}{A^0} \frac{\Delta t}{\Delta x} \right| \leq \frac{1}{3} \leq 1$$

which satisfies the Courant–Friedrich–Lewy (CFL) stability criterion.

Remark 5. A remarkable point is that in this section, we have also proposed a strategy for solving the system (10) at any time. For example at $t = t_1$, first $\mathbf{h}^{1,r}$ is obtained with respect to h_M^1 and Q_{M-1}^1, by solving the system (14). Second, by substituting these values in (13)(b), $\mathbf{Q}^{1,r}$ is computed with respect to h_M^1 and Q_{M-1}^1, too. Now, by (12), h_M^1 and Q_{M-1}^1 are determined and finally the computed values of $\mathbf{h}^1$ and $\mathbf{Q}^1$ are obtained. Similarly, system (10) is solved for the next time.

Remark 6. There is another way in which the oscillatory behavior of the finite element method is minimized, that is, the pairs of spaces V_a and H_a should satisfy the inf–sup condition, i.e.

$$\sup_{h_a \in H_a} \frac{(\nabla . h_a, Q_a)}{\|\nabla h_a\|} \geq \alpha \|Q_a\|_{/}, \quad \forall Q_a \in V_a,$$

where $0 < \|.\|_{/}$ represents the quotient norm (see for instance [14,15]).

4. Numerical experiments

We solve the Saint-Venant equations for a rectangular channel with subcritical and critical flow. Both examples are coded using MATLAB software.

Since $\int_\Omega (\frac{gn^2|Q(x,t_n)|}{R^{4/3}(x,t_n)A(x,t_n)}) Q(x, t_{n+1}) v \mathrm{d}x$ was not computable by MATLAB, and even numerical methods like Simpson's integration rule lead to some complex values, we have used the following simplification:

$$\frac{1}{A(x, t_n)} = \frac{\overline{h}}{\overline{h}A(x, t_n)} = \frac{\overline{h}}{1 - (1 - \overline{h}A(x, t_n))} \approx \overline{h} \sum_{i=0}^{2} (1 - \overline{h}A(x, t_n))^i,$$

where $0 < \overline{h} < 2/\overline{A}$, and $\overline{A}$ is greater than the maximum value of the flow area.

Example 1 (*Subcritical Flow*)**.** The task of estimating the movement of a surge (or shock) or a dam-break wave, resulting from the sudden upstream opening (or the sudden downstream closure) of a sluice gate, in emergencies or dam failures, has occupied the attention of researchers as well as practising engineers for several decades. The determination of the surge height at different locations along the channel provides important information for the design of the bank height. The dreadful disasters due to dam-break flood waves focus the attention of decision-makers on dam-safety problems. We consider an open channel with rectangular cross section whose bottom width is 6.1 m. The bottom slope is 0.00008, the Manning coefficient $n = 0.013$ and the length of the channel is 20 m. The initial conditions in the channel are 5.79 m depth and a steady discharge of 126 m^3/s. The water surface level in the reservoir is constant at the upstream end and also the sluice gate at the downstream end of the channel is suddenly closed at time $t = 0$. We also solved this problem by the finite difference method. Fig. 2 shows the flow depth in the channel at time $t = 0.5$ s obtained by finite element and finite difference methods. As we see, the finite element and finite difference methods are in good agreement. Fig. 3 shows the flow depth at several Δx values for $\Delta t = 1/60$ s and Fig. 4 the flow depth at several Δt values for $\Delta x = 0.33$ with the averaging rule applied for the rectangular channel at time $t = 0.5$ s. If the shape function related to h is not used for the node $x_0 = 0$, then we will observe a large fall in the first intervals (Fig. 5).

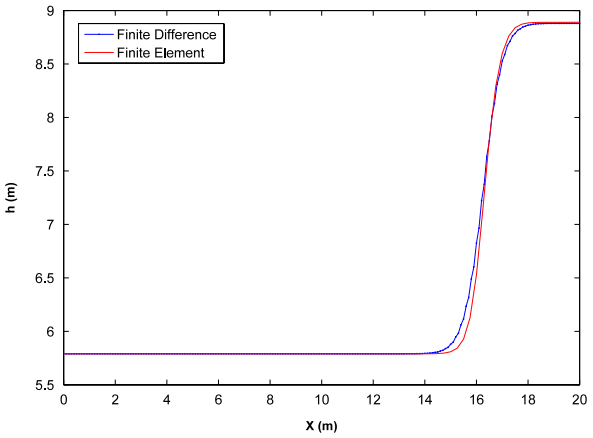

Fig. 2. Flow depth in the rectangular channel at time $t = 0.5$ s obtained by the finite element and finite difference methods.

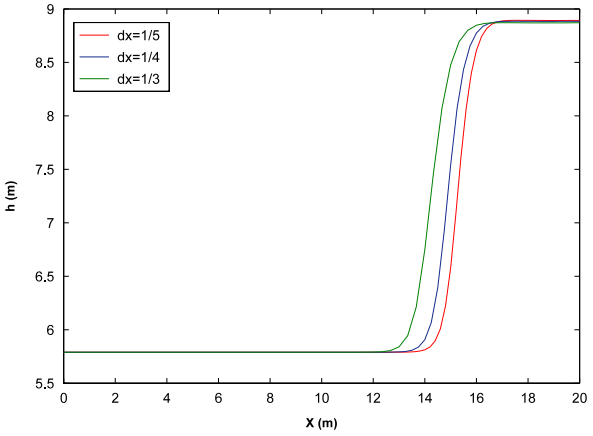

Fig. 3. Flow depth for several Δx values for $\Delta t = 1/60$ s with the averaging rule applied at time $t = 0.5$ s.

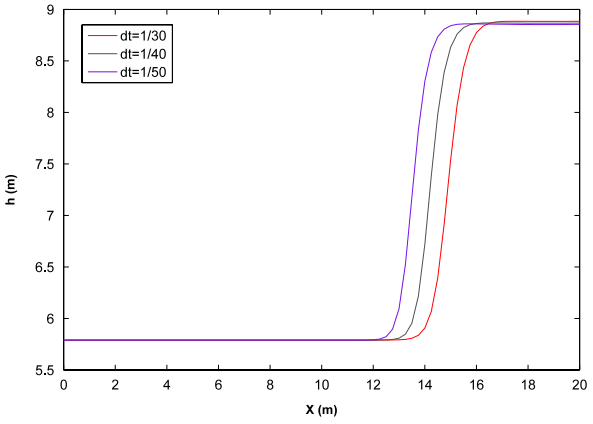

Fig. 4. Flow depth at several Δt values for $\Delta x = 0.33$ with the averaging rule applied at time $t = 0.5$ s.

Example 2 (*Critical Flow*)**.** We consider an open rectangular channel having a bottom width of 6.1 m carrying a flow of 126 m^3/s. The bottom slope is 0.04, the Manning coefficient $n = 0.013$ and the channel length is 20 m. We consider the flow depth $h^0 = \sqrt[3]{\frac{(Q^0)^2}{B^2 g}} = 6.5949$ m. Then $F_r = 1$ and hence there exists a critical flow at $t = 0$. Now, by (28), we take $\frac{\Delta x}{\Delta t} = 2\sqrt{gh^0} = 16.0868$. There is constant level reservoir at the upstream end of the channel. Suppose that a sluice gate at

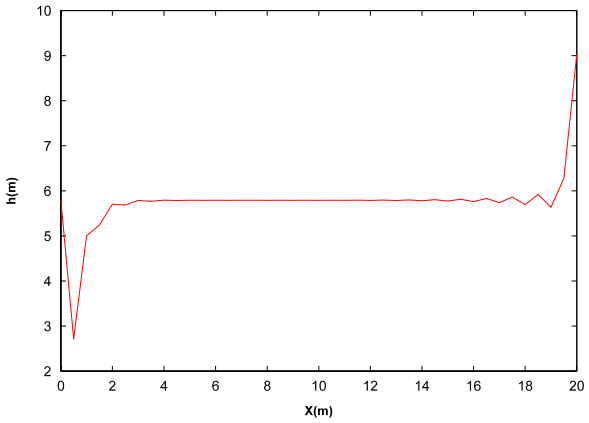

Fig. 5. Flow depth in the rectangular channel at the first time without a shape function related to h for the node $x_0 = 0$.

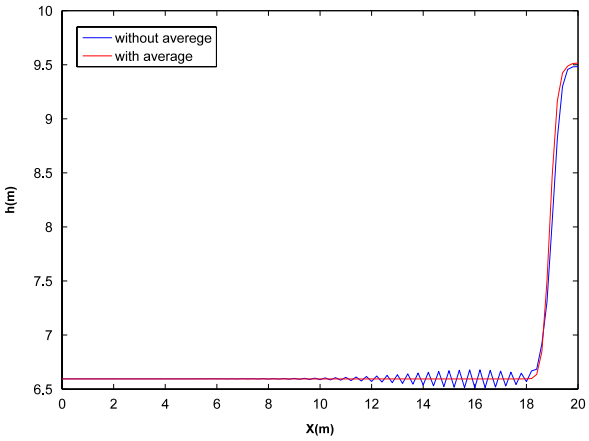

Fig. 6. Flow depth in the rectangular channel at time $t = 0.3$ s with and without the averaging rule applied.

the downstream end is suddenly closed at time $t = 0$. Fig. 6 shows the flow depth in the channel at time $t = 0.3$ s with and without the averaging rule applied, respectively.

5. Conclusion

In this paper, the Saint-Venant equations are discretized by the finite element method such that unwanted oscillations are eliminated (without using an artificial viscosity). Moreover, the numerical stability of the solution is numerically and theoretically investigated, and it is proved that this problem is stable under a simple condition. In future work, discussion on the stability of Saint-Venant equations could be extended to two-dimensional or three-dimensional problems.

References

[1] M.H. Chaudhry, Open-channel Flow, 2nd ed., Springer, 2008.
[2] A.O. Akan, Open Channel Hydraulics, Elsevier, Oxford, UK, 2006.
[3] V.T. Chow, Open Channel Hydraulics, McGraw-Hill, New York, 1959.
[4] M. Amein, H.L. Chu, Implicit numerical modeling of unsteady flows, ASCE J. Hydr. Eng. 101 (HY6) (1974) 717–731.
[5] F.E. Hicks, Finite element modeling of open channel flow, University of Alberta, Ph.D. Thesis, 1990.
[6] R. Szymkiewicz, Finite element method for the solution of the Saint-Venant equation in the open channel network, J. Hydrol. 122 (1991) 275–287.
[7] E. Audusse, M.O. Bristeau, Finite-volume solvers for a multilayer Saint-Venant system, Int. J. Appl. Math. Comput. Sci. 17 (3) (2007) 311–320.
[8] R. Kiladze, Study of the stability of finite difference schemes to solve Saint-Venant equations, Bull. Georgian Natl. Acad. Sci. 3 (1) (2009) 96–99.
[9] G. Bastin, J.M. Coron, B. Andrea-Novel, On Lyapunov stability of linearised Saint-Venant equations for a sloping channel, Netw. Heterog. Media 4 (2) (2009) 177–187.
[10] O. Thual, L.R. Plumerault, D. Astruc, Linear stability of the 1D Saint-Venant equations and drag parameterizations, J. Hydraul. Res. 48 (3) (2010) 348–353.
[11] J. Granatowicz, R. Szymkiewicz, Comparison of efficiency of the solution of the Saint-Venant equations by finite element method and finite difference method, Arch. Hydr. 3–4 (1989) 199–210.
[12] G. Aguilar, F. Gaspar, F. Lisbona, C. Rodrigo, Numerical stabilization of Biot's consolidation model by a perturbation on the flow equation, Int. J. Numer. Methods Eng. 75 (2008) 1282–1300.

[13] Y. Saad, Iterative Methods for Sparse Linear System, 2nd ed., Society for Industrial and Applied Mathematics, 2003.
[14] M.A. Murad, A.F.D. Loula, On stability and convergence of finite element approximations of Biot's consolidation problem, Int. J. Numer. Methods 37 (1994) 645–667.
[15] M.A. Olshanskii, A. Reusken, A Stokes interface problem: stability, finite element analysis and a robust solver, Eur. Cong. Comput. Methods Appl. Sci. Eng. (2004).